

A knowledge-based prognostics framework for railway track geometry degradation

Juan Chiachío^{a,*}, Manuel Chiachío^{a,b}, Darren Prescott^a, John Andrews^a

^a Resilience Engineering Research Group, University of Nottingham, Nottingham NG7 2RD, UK

^b Department of Structural Mechanics and Hydraulics Engineering, University of Granada, Spain



ARTICLE INFO

Keywords:

Railway track degradation
Physics-based modelling
Prognostics
Particle filtering

ABSTRACT

This paper proposes a paradigm shift to the problem of infrastructure asset management modelling by focusing towards forecasting the future condition of the assets instead of using empirical modelling approaches based on historical data. The proposed prognostics methodology is general but, in this paper, it is applied to the particular problem of railway track geometry deterioration due to its important implications in the safety and the maintenance costs of the overall infrastructure. As a key contribution, a knowledge-based prognostics approach is developed by fusing on-line data for track settlement with a physics-based model for track degradation within a filtering-based prognostics algorithm. The suitability of the proposed methodology is demonstrated and discussed in a case study using published data taken from a laboratory simulation of railway track settlement under cyclic loads, carried out at the University of Nottingham (UK). The results show that the proposed methodology is able to provide accurate predictions of the *remaining useful life* of the system after a model training period of about 10% of the process lifespan.

1. Introduction

In most developed countries, the continuous ageing and the growing demand of use of critical infrastructures calls for advanced Prognostics and Health Management (PHM) concepts for optimal infrastructure asset management [1]. In particular for the railway infrastructure, the continuous deterioration of the track due to traffic loading represents the main ageing factor requiring periodic interventions to restore the track to an acceptable geometry [2,3]. These maintenance interventions not only represent a significant part of the railway operation expenses, but also imply temporary line closures and a reduction of the effective network capacity, so they need to be planned months in advance. It is in this context of anticipated decision making where the benefits of PHM can be fully exploited for maintenance cost optimization and downtime reduction.

As evident from the literature, railway track degradation and maintenance modelling to date has a strong empirical retrospective character, mainly grounded on data-based (stochastic or phenomenological) models with limited prospective capability. A systematic review of these models is proposed by Dahlberg [4] for track settlement modelling, and more recently by Soleimanmeigouni *et al.* [5] and Higgins and Liu [6] focusing also in track maintenance modelling. The prediction accuracy of those models strongly depends on the quality

and quantity of the available historic data, and thus they are prone to misjudgements especially under medium-to-long term future scenarios. To overcome this limitation, some physics-based models have been proposed to predict the progressive mechanical degradation of the track under cyclic loadings from first geomechanical principles, like the cyclic densification model by Suiker and de Borst [7], and the elastoplastic models by Indraratna *et al.* [8] and more recently by Sun *et al.* [9]. The referred physics-based models are transparent to human understanding and they can adapt to different material and loading conditions without much training, which is essential to confer the required early predictive capability. However, they are unable to account for the uncertainty in the predictions since they are based on deterministic input-output relationships. In this context, a number of authors have adopted statistical methods to quantify the modelling uncertainty in track geometry degradation, such as regression analysis [10], stochastic processes [11–13], Markov models [14,15], and Bayesian analysis [16], to cite but a few. Notwithstanding, as stated before, the accuracy of these methods relies almost entirely on historical data, which bounds their prospective capability and the reliability of their predictions.

In this paper, a paradigm shift for track geometry degradation modelling and maintenance is proposed. Instead of making maintenance decisions based on either a retrospective data-based modelling

* Corresponding author.

E-mail address: juan.chiachioruano@nottingham.ac.uk (J. Chiachío).

<https://doi.org/10.1016/j.ress.2018.07.004>

Received 1 December 2017; Received in revised form 24 June 2018; Accepted 4 July 2018

Available online 05 July 2018

0951-8320/ © 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

of the track, or, alternatively, using a purely deterministic physics-based approach, a *knowledge-based* prognostics framework is proposed herein. This approach fuses information from physics-based models and available data about track degradation within a Bayesian learning paradigm to sequentially reduce the initial modelling uncertainty [17] as long as new data are collected, so as to obtain increasingly accurate forecasts of the future condition of the track. These forecasts enable the testing of various load and utilisation scenarios, and more importantly, allow us to answer the question of “when will failure occur” with quantified uncertainty. The proposed methodology enables informed, anticipated, and risk-based decisions about optimal railway track asset management, and can be extended to other railway assets to enable decision-making at a system infrastructural level. To the authors’ best knowledge, the existing literature on track degradation prognostics is currently limited to Mishra *et al.* [18], focusing on railway switches based on a phenomenological model for track geometry degradation.

For the problem of physical modelling of the track deterioration, an elasto-plastic model for track settlement is developed based on first principles of the plasticity of soils [19] and Critical State Soil Mechanics (CSSM) [20]. This model provides the temporal evolution of the elementary volumetric and deviatoric plastic strains of the track based on some geomechanical inputs. Apart from the traffic loads, the inputs required by the model reduce to the *critical state parameters* [20], which can be obtained from routine triaxial tests, and two uncertain parameters controlling the cyclic hardening of the granular layers of the track, which are sequentially updated using condition monitoring data. For the sake of Bayesian data assimilation, the proposed model is embedded within a hidden Markov model [21] where the deviatoric and volumetric plastic strains constitute the *latent states*, and the vertical permanent axial strain is the *observed state*. Next, a sequential Bayesian updating framework is used to update both the latent states and the uncertain model parameters as long as new data become available. The resulting equations for sequential updating are analytically intractable, hence a Sequential Monte Carlo approach based on *particle filters* [22] is adopted whereby the probability density functions (PDFs) are approximated by weighted samples or *particles*, which represent random realisations of the updated track degradation states. Finally, by extrapolating those particles into the future, a probability-based estimation of the time (load cycles) to reach a predefined functional limit for track degradation is subsequently obtained. This time defines what in the Prognostics and Health Management community is commonly known as the *remaining useful life* (RUL) [1] of the system.

As a case study, the proposed methodology is tested against experimental data taken from Aursudkij *et al.* [23] about permanent axial strain in a ballasted railway track, carried out at the Nottingham Railway Test Facility [24]. The results show that the proposed prognostics methodology is able to accurately anticipate the future states of deformation of the track after a training period of about 10% of the total length of the process, which corresponds to the time required by the model to assimilate the data. This prognostics performance is compared with the performance obtained for the same dataset using an empirical logarithmic model for track settlement. The assessment is carried out using a newly developed prognostics metric named as the *relative prospective evidence*, which measures the relative prediction accuracy of two different model classes in relation to the amount of information that the models need to extract from the data to perform their predictions. The results show that the physics-based model proposed in this paper is the one with the larger and earlier anticipation capacity, needing less support from condition monitoring data to make accurate prognostics estimates.

The remainder of the paper is organised as follows. Section 2 presents the physical fundamentals behind the proposed model for track deterioration. Section 3 overviews the mathematical basis and computational aspects of Bayesian state estimation and prognostics, which have been both specialised for their application to the railway track

degradation problem. In Section 4, the proposed framework is applied to railway track settlement data to serve as a case study. A discussion about the suitability of model-based prognostics for railway track maintenance planning is provided in Section 5. Finally, Section 6 provides concluding remarks.

2. Physics-based model for track settlement

A constitutive model for ballast plastic deformation under cycling loading is developed here following the general theory of plasticity, CSSM, and Pender’s postulates about plastic deformation of over-consolidated soils [25]. The model predicts the evolution of the plastic vertical strain ϵ_v^p of a representative track section, which is the variable of interest here, whereby the settlement of the section can be straightforwardly derived. The reader is referred to [25] for a detailed overview of the foundations of the proposed model, but for the sake of clarity, the key formulation and definitions are reproduced here with a uniform notation.

2.1. Physical fundamentals

Following the theory of plasticity, a granular medium under a stress state represented by the stress tensor σ_{ij} , will experience an increment in the plastic strains given by Hill’s *flow rule* [26], as

$$d\epsilon_{ij}^p = h \frac{\partial g}{\partial \sigma_{ij}} df \quad (1)$$

where $d\epsilon_{ij}^p$ is the corresponding plastic strain increment tensor, df is the differential of function f , which denotes the *yield criterion* (or *yield surface*), g is the *plastic potential*, and h is the *hardening function*. These elements are briefly explained in Definitions 1 to 3, respectively. If the stress invariants p and q are adopted instead of the stress tensor σ_{ij} , Eq. (1) can be rewritten as:

$$d\epsilon_v^p = h \frac{\partial g}{\partial p} df \quad (2a)$$

$$d\epsilon_s^p = h \frac{\partial g}{\partial q} df \quad (2b)$$

with ϵ_v^p and ϵ_s^p being the corresponding plastic volumetric and distortional strain invariants. The reader is referred to Appendix A for further insight about stress and strains invariants. Having ϵ_v^p and ϵ_s^p as known, the vertical plastic strain of the track section can be straightforwardly computed as:

$$\epsilon_v^p = 1/3\epsilon_v^p + \epsilon_s^p \quad (3)$$

as shown in Eq. (A.4b) in Appendix A. It follows that functions f , g and h need to be determined so as to define a closed-form expression for the constitutive model in Eq. (2).

Definition 1 (Yield criterion [19]). *The yield criterion defines the limit of the elastic behaviour and the onset of plastic deformation under any possible combination of stresses. Mathematically: $f(p, q) = 0$.*

Assuming Pender’s hypothesis that overconsolidated soils experience plastic strain only when there is a change in the stress ratio ($\eta = q/p$) [25], the yield criterion can be assumed to be given by:

$$f = q - \eta_i p = 0 \quad (4)$$

where η_i defines a particular constant stress ratio yield locus i (refer to Fig. 2 in [25]). By differentiating Eq. (4), $df = dq - \eta_i dp$, and substituting $dq = \eta_i dp + p d\eta$ obtained from the definition of η , the required expression for df in Eq. (2) can be shown to be given by:

$$df = p d\eta \quad (5)$$

Definition 2 (Plastic potential [19]). *The plastic potential is given by the*

generic expression $g(p, q) = 0$ and relates the incremental plastic strains with the stress state following the plastic flow rule in Eq. (1).

An analytical expression for g can be obtained by noting that the plastic strains are normal to the plastic potential $g(p, q) = 0$, thus implying that [19]:

$$\frac{d\epsilon_v^p}{d\epsilon_s^p} = -\frac{dq}{dp} \quad (6)$$

In this research, the ratio $d\epsilon_v^p/d\epsilon_s^p$ is assumed to be given by a modified version of Rowe's stress-dilatancy model [27], as

$$\frac{d\epsilon_v^p}{d\epsilon_s^p} = \frac{9(M - \eta^*)}{9 + 3M - 2M\eta^*} \quad (7)$$

where M is the slope of the critical state line in the $p - q$ space [28], and η^* is a function of the stress ratio given by $\eta^* = \eta p/p_{cs}$, with p_{cs} being defined in Eq. (13) below. The stress-dilatancy criterion in Eq. (7) differs from that proposed by Rowe [27] in that it provides positive values for plastic volumetric strain increments (i.e., contraction), as long as η^* does not exceed the critical state value M . By substituting Eq. (7) into (6) and solving the resulting differential equation, an expression for the plastic potential function is obtained, as:

$$g(p, q) = 3M \ln(p/C) + (2M + 3)\ln(2\eta^* + 3) - (3 - M)\ln(3 - \eta^*) = 0 \quad (8)$$

where C is an integration constant.

Definition 3 (Hardening function [19]). As a work hardening material, ballast experiences a dynamical change of its plastic response during yielding. This change is accomplished by modifying the yield surface as plastic flow occurs. The hardening function (h in Eq. (2)) accounts for such effect.

An expression for h can be obtained following the premise that no volume change takes place during undrained deformation [20]. Mathematically:

$$d\epsilon_v = d\epsilon_v^e + d\epsilon_v^p = 0 \quad (9)$$

where $d\epsilon_v^e$ is the elastic volumetric strain increment given by [20]

$$d\epsilon_v^e = \frac{\kappa dp}{p(1 + e_0)} \quad (10)$$

with κ being the swelling/recompression constant. By substituting Eqs. (2a) and (10) into Eq. (9), h is obtained as:

$$h = \frac{-\kappa}{1 + e_0} \frac{dp}{p} \left(\frac{\partial g}{\partial p} p d\eta \right)^{-1} \quad (11)$$

In Eq. (11), Pender's assumption of parabolic undrained stress paths in the $p - q$ space [25] is adopted to relate the stress derivatives dp and $d\eta$. This assumption is expressed as:

$$\left(\frac{\eta - \eta_0}{M - \eta_0} \right)^2 = \frac{p_{cs}}{p} \left(\frac{1 - p_0/p}{1 - p_0/p_{cs}} \right) \quad (12)$$

where η_0 is the initial stress ratio (corresponding to the minimum load within the cycle), p_0 is the initial value of the mean stress invariant p , and p_{cs} is the value of p at the point on the critical state line corresponding to the current voids ratio e , obtained as [28,29]:

$$p_{cs} = \exp\left(\frac{\Gamma - e}{\lambda_{cs}}\right) \quad (13)$$

with Γ and λ_{cs} being material parameters. These parameters, together with M and κ (see nomenclature Table A.3), are the critical state parameters of the model and can be evaluated from a series of drained triaxial compression tests [30]. Having obtained the differential dp using Eq. (12), and substituting it into Eq. (11), the hardening function

becomes after some algebraic manipulation¹:

$$h = \frac{2\kappa(1 - p_0/p_{cs})(p/p_{cs})(\eta - \eta_0)}{(1 + e_0)(M - \eta_0)^2(2p_0/p - 1)} \frac{9 + 3M - 2M\eta^*}{9(M - \eta^*)} d\eta \quad (14)$$

It should be emphasised that the last equation is nominally valid for the first load cycle. For subsequent loading, h should account for other complex phenomena observed in experiments under cyclic loading conditions, such as the Bauschinger effect, the effect of the stress ratio and loading history, etc. [19]. A semi-empirical approach is adopted here to account for such processes following Indraratna and Salim [8], which essentially consists of multiplying the hardening function in Eq. (14) by a semi-empirical factor ϕ given by

$$\phi = \left(1 - \frac{\eta^*}{M}\right) \left(\frac{\langle p - p_e \rangle^2 + \langle q - q_e \rangle^2}{(p_{max} - p_{min})^2 + (q_{max} - q_{min})^2} \right)^{1/2} \alpha n^{-\beta} \quad (15)$$

where α and β are empirical fitting parameters, n is the accumulated number of cycles, $\langle \cdot \rangle$ are the Macauley brackets, and p_e is given by the expression [8]:

$$p_e = p_{min} + \left(1 - \frac{1}{\ln(n + 10)}\right)(p_{max} - p_{min}) \quad (16)$$

with p_{min} and p_{max} being the minimum and maximum mean stress invariant within load cycle n , respectively. The corresponding deviatoric components q_{min} , q_{max} and q_e are obtained using Eq. (A.4a), defined in Appendix A.

2.2. Constitutive equations of proposed model

Substituting the expressions for f , g and h obtained above into Eq. (2), and after some algebraic manipulation, the following expressions for the evolution of the plastic volumetric and deviatoric strains are obtained:

$$\frac{d\epsilon_v^p}{d\eta} \bigg|_n - \frac{2\phi\kappa(1 - p_0/p_{cs})(p/p_{cs})(\eta - \eta_0)}{(1 + e_0)(M - \eta_0)^2(2p_0/p - 1)} \bigg|_n = 0, \quad \epsilon_v^p(\eta_0)|_n = \epsilon_v^p(\eta_{max})|_{n-1} \quad (17a)$$

$$\frac{d\epsilon_s^p}{d\eta} - \frac{9 + 3M - 2M\eta^*}{9(M - \eta^*)} \frac{d\epsilon_v^p}{d\eta} \bigg|_n = 0, \quad \epsilon_s^p(\eta_0)|_n = \epsilon_s^p(\eta_{max})|_{n-1} \quad (17b)$$

where $\cdot|_n$ denotes a variable evaluated at load cycle n , and $\eta_{max} = q_{max}/p_{max}$. Note that for the n th loading cycle, the initial condition for both deviatoric and volumetric plastic strains is given by the corresponding plastic strains at the previous cycle $n - 1$. The initial conditions $\epsilon_v^p(\eta_0)|_0$ and $\epsilon_s^p(\eta_0)|_0$ are assumed as known. Observe also that the expressions in Eq. (17) are non-linear differential equations of the plastic strains, thus they need to be integrated numerically by replacing the derivatives by their finite difference approximation (e.g., using explicit Euler method). An algorithmic description of the numerical method proposed to integrate the constitutive Eqs. (17a) and (17b) is presented in Algorithm 1.

3. Track-degradation prognostics

Prognostics is concerned with predicting the future state of health of engineering systems or components given the current degree of wear or damage, and, based on that, estimating the remaining time in which the system is expected to continue performing its intended function within desired specifications [1]. The predictions are carried out using no additional evidence but a sequence of measurements up to the time of prediction, and the available models for system behaviour. A physics-based filtering-based prognostics framework is proposed herein as depicted in Fig. 1. The definitions in Section 3.1 are used to provide the

¹ In the normal range of stresses, ballast remains in states denser than the critical state [29], thus, the sign of the hardening function in Eq. (14) is reversed.

Input: σ_3 {Confining pressure}; $[M, \lambda_{cs}, \Gamma, \kappa]$ {critical state parameters}; $[\alpha, \beta]$ {fitting parameters}; $[\sigma_{1,min}, \sigma_{1,max}]$ {min, max vertical stress}; n {# of load cycles}; N_S {# of in-cycle load steps}; ϵ_{v0}^p {initial elastic volumetric strains}; e_0 {initial voids ratio}; $[\epsilon_{v,n}^p, \epsilon_s^p]_n$ {plastic strains after n cycles}; e_n {voids ratio after n cycles}

Output: $[\epsilon_{v,n+1}^p, \epsilon_s^p]_{n+1}$ {plastic strains at $n+1$ }; e_{n+1} {voids ratio at $n+1$ }

Begin:

- 1: Set min and max deviatoric stresses: $[q_{min}, q_{max}]^T = [\sigma_{1,min} - \sigma_3, |\sigma_{1,max} - \sigma_3|]$
- 2: Set min and max mean stresses: $[p_{min}, p_{max}]^T = 1/3[q_{min}, q_{max}]^T + \sigma_3[1, 1]^T$
- 3: Set in-cycle stress increments $[\Delta p, \Delta q]^T = ([p_{max}, q_{max}]^T - [p_{min}, q_{min}]^T)/N_S$
- 4: Set initial stress state $[q_0, p_0] = [q_{min}, p_{min}]$
- 5: Set initial voids ratio: $e_{i=1} \leftarrow e_n$
- 6: **for** $i = 1$ to N_S **do**
- 7: Obtain $(p_{cs})_i = p_{cs}(e_i)$ from Eq. (13)
- 8: Obtain $(p_0)_i = p_0(q_i, p_i, (p_{cs})_i)$ from Eq. (12)
- 9: Compute stress state: $[p_i, q_i]^T = [p_{i-1}, q_{i-1}]^T + [\Delta p, \Delta q]^T$
- 10: Set stress ratio increment: $\Delta \eta_i = \frac{q_i}{p_i} - \frac{q_{i-1}}{p_{i-1}}$
- 11: Obtain $(d\epsilon_v^e/dp)_i$ from Eq. (10)
- 12: Compute volumetric elastic strain: $(\epsilon_v^e)_i = (\epsilon_v^e)_{i-1} + (d\epsilon_v^e/dp)_i \Delta p$
- 13: Compute $[d\epsilon_v^p/d\eta]_i, [d\epsilon_s^p/d\eta]_i$ from Eqs. (17a) and (17b)
- 14: Compute plastic strains: $[(\epsilon_v^p)_i, (\epsilon_s^p)_i]^T = [(\epsilon_v^p)_{i-1}, (\epsilon_s^p)_{i-1}]^T + [(d\epsilon_v^p/d\eta)_i, (d\epsilon_s^p/d\eta)_i] \Delta \eta_i [1, 1]^T$
- 15: Obtain total volumetric strain: $(\epsilon_v)_i = (\epsilon_v^p)_i + (\epsilon_v^e)_i$
- 16: Update the voids ratio $e_{i+1} = e_0 - (1 + e_0)(\epsilon_v)_i$
- 17: **end for**
- 18: $[e_{n+1}, \epsilon_{v,n+1}^p, \epsilon_s^p]_{n+1} \leftarrow [e_{N_S}, (\epsilon_v^p)_{N_S}, (\epsilon_s^p)_{N_S}]$

Algorithm 1. Step-by-step description of model for track settlement.

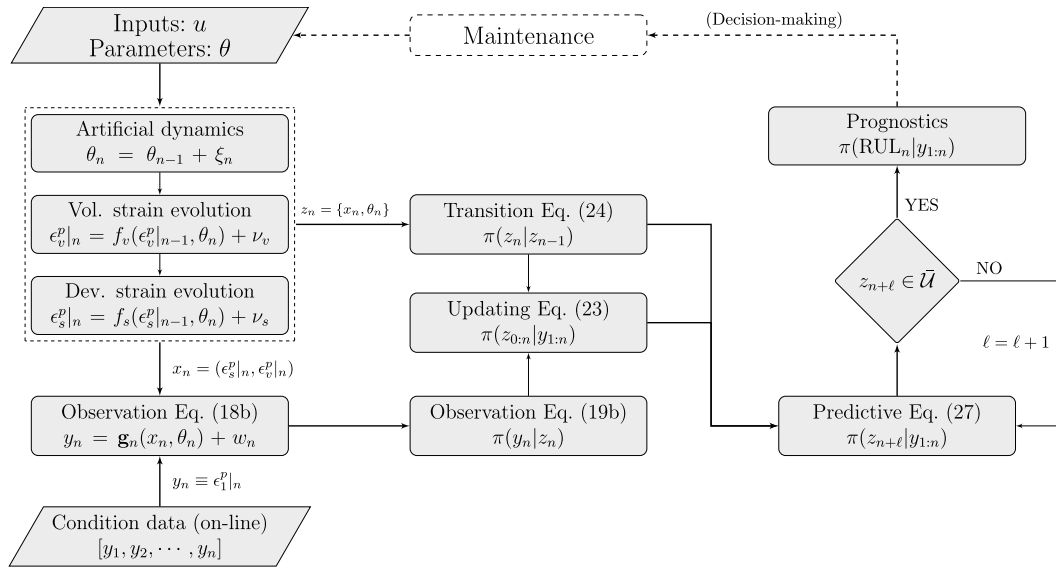


Fig. 1. Schematic overview of the proposed prognostics framework.

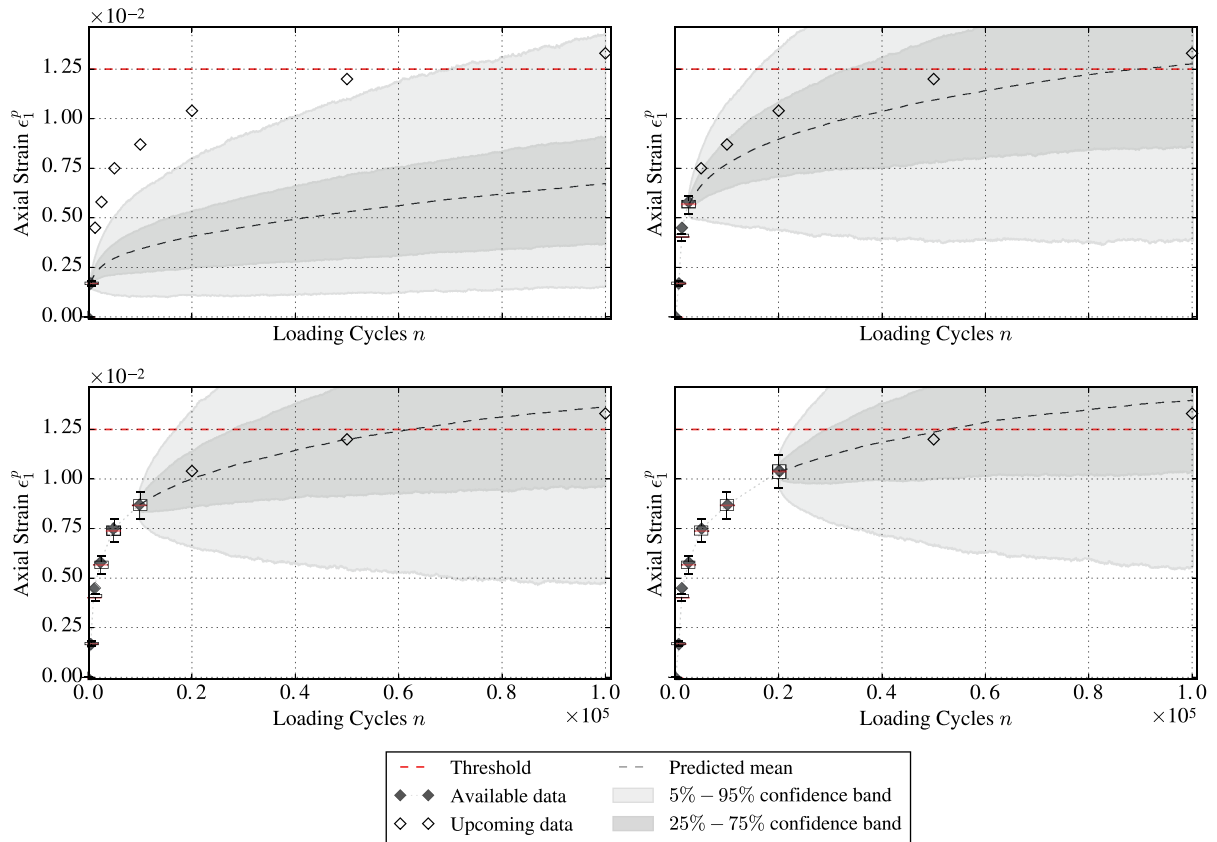


Fig. 2. Sequential estimation and future prediction of permanent axial strain at 600 (upper left), 2000 (upper right), 10,000 (lower left) and 20,000 (lower right) loading cycles. The updated state estimations up to prediction time n are represented using a box-and-whisker plot, where the bottom and top of the box represent the 25th and 75th percentiles, and the lower and upper whiskers the 5th and 95th percentiles, respectively. The future states prediction are represented using shaded areas for the 5% – 95% and 25% – 75% probability bands.

required mathematical background about prognostics in the context of the track degradation problem investigated here.

3.1. Fundamentals about model-based prognostics

Definition 4 (Hidden Markov model [21]). A hidden Markov model is a

time-series model comprised of a couple of discrete-time stochastic processes, namely $\{x_n\}_{n \geq 0}$ and $\{y_n\}_{n \geq 0}$. The former is an unobserved Markov process described by the state vector x_n taking values in a space denoted by $\mathcal{X} \subset \mathbb{R}^{n_x}$, with known prior PDF $\pi(x_0)$ and transition density given by $\pi(x_n|x_{n-1})$, non necessarily Gaussian. The process $\{y_n\}_{n \geq 0}$ represents a sequence of noisy observations taking values in $\mathcal{Y} \subset \mathbb{R}^{n_y}$, which are assumed

to be conditionally independent given the states x_n .

A discrete-time state-space representation is typically adopted to describe the hidden Markov model, as follows [1]:

$$x_n = f_n(x_{n-1}, \theta, u_n) + v_n \quad (18a)$$

$$y_n = g_n(x_n, \theta, u_n) + w_n \quad (18b)$$

where $f_n: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}^{n_x}$ is the *state transition equation*, represented by a possibly nonlinear function of the state variable x_n along with a set $u \in \mathbb{R}^{n_u}$ of input parameters to the system (loadings, geo-mechanical inputs, etc.). The function $g_n: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}^{n_y}$ denotes the *observation equation* and provides the expected observation or measurement y_n given the hidden state x_n . Both functions f and g may depend on a set of uncertain model parameters $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$ (e.g., fitting parameters). The terms $v_n \in \mathbb{R}^{n_x}$ and $w_n \in \mathbb{R}^{n_y}$ in Eq. (18) are stochastic variables that represent the modelling error and the measurement error, respectively. Supported by the Principle of Maximum Information Entropy [31], v_n and w_n can be conservatively modelled as zero-mean Gaussian distributions [1]. Thus, the state space model defined in Eq. (18) can be probabilistically rewritten as²:

$$\pi(x_n | x_{n-1}, \theta) = \mathcal{N}(f_n(x_{n-1}, \theta), \Sigma_v) \quad (19a)$$

$$\pi(y_n | x_n, \theta) = \mathcal{N}(g_n(x_n, \theta), \Sigma_w) \quad (19b)$$

where $\Sigma_v \in \mathbb{R}^{n_x \times n_x}$ and $\Sigma_w \in \mathbb{R}^{n_y \times n_y}$ are the covariance matrices of the model error and the measurement error, respectively. Note in Eq. (19) and in what follows that the conditioning on the model inputs u_n is dropped for simpler notation.

In this work, the (latent) state of the system at loading cycle n is assumed to be described by the plastic deviatoric and volumetric strains, i.e., $x_n = (\epsilon_s^p | n, \epsilon_v^p | n) \in \mathbb{R}^2$, where subscripts (s, v) denote the deviatoric and volumetric components, respectively. The observations of the system are assumed to be given by the vertical plastic strain, $y_n = \epsilon_v^p | n \in \mathbb{R}$, which can be theoretically estimated from the component plastic strains using Eq. (3). Under these assumptions, the covariance matrices for the error terms are given by $\Sigma_v = \text{diag}(\sigma_{v,s}, \sigma_{v,v}) \in \mathbb{R}^{2 \times 2}$ for the model error, and $\Sigma_w = \sigma_w \in \mathbb{R}$ for the measurement error, with $\sigma_{v,s}$ and $\sigma_{v,v}$ being respectively the standard deviations of the deviatoric and volumetric components of the model error, and σ_w the standard deviation of the measurement error. With these settings, the deterministic physics-based model proposed in Section 2 is *stochastically embedded* within a hidden Markov model by defining the functions f and g in Eqs. (19a) and (19b), as:

$$f_n(x_{n-1}, \theta) = (f_s(\epsilon_s^p | n-1, \theta), f_v(\epsilon_v^p | n-1, \theta)) \quad (20a)$$

$$g_n = \epsilon_s^p | n + \frac{1}{3} \epsilon_v^p | n \quad (20b)$$

where functions f_s and f_v in Eq. (20a) are given by

$$f_s(\epsilon_s^p | n-1, \theta) = \epsilon_s^p | n-1 + \Delta \epsilon_s^p | n(\theta) \quad (21a)$$

$$f_v(\epsilon_v^p | n-1, \theta) = \epsilon_v^p | n-1 + \Delta \epsilon_v^p | n(\theta) \quad (21b)$$

In the last equation, $\Delta \epsilon_s^p | n$ and $\Delta \epsilon_v^p | n$ are the increments in the plastic deviatoric and volumetric strains within loading cycle n , respectively. These strain increments can be obtained as:

$$\Delta \epsilon_{s,v}^p | n = \sum_{j=1}^{N_G} \left(\frac{d\epsilon_{s,v}^p}{d\eta} \right)_{\eta_j} \Delta \eta_j \quad (22)$$

² In what follows, the probability density functions are denoted as $\pi(\cdot)$ instead of $p(\cdot)$, to avoid confusion with the mean stress invariant p .

where subscripts (s, v) denote the deviatoric and volumetric components, respectively, and $N_G \in \mathbb{N}$ is the number of discretising steps of the loading ramp at cycle n (the unloading ramp is regarded as elastic). The reader is referred to Algorithm 1 for details about the numerical computation of the plastic deviatoric and volumetric strains for load cycle n .

Definition 5 (Sequential state estimation). The sequential estimation of the track degradation aims at obtaining an updated estimation of the actual condition of the track based on available measurements up to cycle n . The degradation is described by the state variable x_n along with the uncertain model parameters θ , leading to an augmented state variable $z_n = (x_n, \theta) \in \mathcal{Z} \subset \mathbb{R}^{n_x+n_\theta}$. Thus, given a sequence of track degradation measurements, denoted as $y_{1:n} \triangleq \{y_1, \dots, y_{n-1}, y_n\}$, the goal is to obtain the most plausible sequence of track degradation states $z_{0:n} \triangleq \{z_0, \dots, z_{n-1}, z_n\}$ for those measurements, which is given by Bayes' Theorem as [1]:

$$\begin{aligned} \pi(z_{0:n} | y_{1:n}) &= \frac{\pi(y_n | z_n) \pi(z_{0:n} | y_{1:n-1})}{\int_{\mathcal{Z}} \pi(y_n | z_n) \pi(z_{0:n} | y_{1:n-1}) dz_{0:n}} \\ &\propto \underbrace{\pi(y_n | z_n)}_{\text{Eq. (19b)}} \underbrace{\pi(z_n | z_{n-1}) \pi(z_{0:n-1} | y_{1:n-1})}_{\text{last update}} \end{aligned} \quad (23)$$

where

$$\pi(z_n | z_{n-1}) = \underbrace{\pi(x_n | x_{n-1}, \theta_n)}_{\text{Eq. (19a)}} \pi(\theta_n | \theta_{n-1}) \quad (24)$$

Observe from Eq. (24) that model parameters θ_n are virtually time-evolving although they are essentially not dependent on time, which is a key problem that typically arises when sequentially updating the state variable as an augmented state [32]. A common solution is to add a white-noise perturbation to the set of updated parameters at time $n-1$ before evolving to the next predicted state at time n , i.e., $\theta_n = \theta_{n-1} + \xi_n$, where $\xi_n \sim \mathcal{N}(0, \Sigma_{\xi_n})$. This induces a Markovian-type artificial dynamics to the model parameters whereby the required PDF $\pi(\theta_n | \theta_{n-1})$ is prescribed [32,33], as follows:

$$\pi(\theta_n | \theta_{n-1}) = \mathcal{N}(\theta_{n-1}, \Sigma_{\xi_n}) \quad (25)$$

where is a covariance matrix specified here as $\Sigma_{\xi_n} = \text{diag}(\sigma_{\xi_{n,1}}^2, \dots, \sigma_{\xi_{n,j}}^2, \dots, \sigma_{\xi_{n,n_\theta}}^2)$, with $\sigma_{\xi_{n,j}}^2$ being the variance of the random walk of the j th component of the parameter vector θ at time or cycle n . Note that such artificial evolution imposes a loss of information in θ over time (e.g., increasingly larger spread in $\pi(z_{0:n} | y_{1:n})$) since additional uncertainties are artificially added to the model parameters. Several methods have been proposed in the literature to overcome this drawback, with the most popular being those that impose a shrinkage over Σ_{ξ_n} as long as new data are available [32]. An efficient method of this class is adopted in this work [34], which consists in sequentially modifying the variances $\sigma_{\xi_{n,j}}^2$, $j = 1, \dots, n_\theta$, as follows:

$$\sigma_{\xi_{n,j}}^2 = \sigma_{\xi_{n-1,j}}^2 \left(1 - P_j^* \frac{\text{RMAD}(\theta_{n,j}) - \text{RMAD}_j^*}{\text{RMAD}(\theta_{n,j})} \right) \quad (26)$$

where $\text{RMAD}(\theta_{n,j})$ is the relative median absolute deviation of $\pi(\theta_{n,j} | y_{1:n})$, RMAD_j^* is the target RMAD for $\pi(\theta_{n,j} | y_{1:n})$, and $P_j^* \in [0, 1]$ is a scaling constant that controls the speed of convergence to RMAD_j^* . Both RMAD_j^* and P_j^* need to be tuned by the modeller. A comprehensive discussion about the optimal choice for P_j^* and RMAD_j^* is found in [34].

Definition 6 (Predicted system state). Given the most up-to-date assessment of the system at time or cycle n , $\pi(z_n | y_{1:n})$, a probability-based prediction of the future states of degradation of the system ℓ -steps forward in time can be obtained by Total Probability Theorem as [1,35]:

$$\pi(z_{n+\ell} | y_{1:n}) = \int_{\mathcal{Z}} \left[\prod_{t=n+1}^{n+\ell} \pi(z_t | z_{t-1}) \right] \pi(z_n | y_{1:n}) dz_{n:n+\ell-1} \quad (27)$$

where $\ell > 1$ is the number of prediction steps in absence of new data, and $\pi(z_t|z_{t-1})$ is the state transition equation defined in Eq. (24), which, for $t > n$, encloses the future behaviour of the system.

To numerically solve this multi-dimensional integral, an approximation can be readily obtained by *conditional sampling*, using recursively the one-step transition equation defined in (24), i.e.: first sample $z_{n+1}^{(i)}$ using the aforementioned one-step transition equation conditional on the initial state z_n , i.e., $z_{n+1}^{(i)} \sim \pi(\cdot|z_n)$; then sample the succeeding state conditional on the previous sample, i.e., $z_{n+2}^{(i)} \sim \pi(\cdot|z_{n+1}^{(i)})$; finally, repeat the same process until the target time $n + \ell$ is reached.

Definition 7 (End of Life (EOL) [1]). The EOL is defined as the earliest time or loading cycle $t > n$ when the predicted event $[z_t \in \overline{\mathcal{U}}]$ takes place, where $z_t \sim \pi(z_t|y_{1:n})$ and $\overline{\mathcal{U}} \subset \mathcal{Z}$ is the subset of states where the system behaviour becomes unacceptable.

In the context of the proposed track degradation problem, $\overline{\mathcal{U}}$ represents the set of values for the vertical plastic strain exceeding the threshold $\partial\overline{\mathcal{U}}$, which delimits the boundary between the *failure region* $\overline{\mathcal{U}}$ and the *safe region* $\mathcal{U} = \mathcal{Z} \setminus \overline{\mathcal{U}}$. Observe that since the predicted states z_t are uncertain (recall Eq. (27)), the EOL will be an uncertain variable distributed according to the PDF $\pi(\text{EOL}_n|y_{1:n})$. This PDF can be theoretically obtained as the distribution of the predicted states contained in the boundary $\partial\overline{\mathcal{U}}$, i.e., $\pi(\text{EOL}_n|y_{1:n}) \equiv \pi(z_t \in \partial\overline{\mathcal{U}}|y_{1:n})$, given that $[t = \text{EOL}_n] \Leftrightarrow [z_t \in \partial\overline{\mathcal{U}}]$. It follows that $\pi(z_{n+\ell} \in \partial\overline{\mathcal{U}}|y_{1:n})$ can be written as a conditional PDF $\pi(z_{n+\ell}|y_{1:n}, \partial\overline{\mathcal{U}})$, which can be evaluated as:

$$\pi(z_{n+\ell}|y_{1:n}, \partial\overline{\mathcal{U}}) = \pi(\text{EOL}_n|y_{1:n}) = \frac{k \pi(z_t|y_{1:n}) \mathbb{I}_{\partial\overline{\mathcal{U}}}(z_t)}{\text{Eq. (27)}} \quad (28)$$

where k is a normalising constant and $\mathbb{I}_{\partial\overline{\mathcal{U}}}(z_t)$ is an indicator function that assigns the unity if $(z_t \in \partial\overline{\mathcal{U}})$, and makes zero the rest.

Definition 8 (Remaining Useful Life (RUL) [1]). The RUL is the remaining time in which the system is foreseen to perform within the region \mathcal{U} of acceptable performance, where $\mathcal{U} = \mathcal{Z} \setminus \overline{\mathcal{U}}$, the non-empty subset of authorised states of the system. A prediction of the RUL from the time of prediction n can be straightforwardly obtained from EOL_n as $\text{RUL}_n = \text{EOL}_n - n$.

Definition 9 (Prospective evidence for a model). Given a degradation process \mathcal{D}_n observed up to current time or load cycle n , and a particular model class \mathcal{M} which idealises such a process (e.g., the one proposed in Section 2), the prospective evidence for model class \mathcal{M} is the probability of the subsequent degradation process, denoted as \mathcal{D}_n^+ , to be predicted by \mathcal{M} . The prospective evidence can be obtained by Total Probability Theorem as:

$$\pi(\mathcal{D}_n^+|\mathcal{D}_n, \mathcal{M}) = \int_{\Theta} \pi(\mathcal{D}_n^+|\theta_n, \mathcal{D}_n, \mathcal{M}) \pi(\theta_n|\mathcal{D}_n, \mathcal{M}) d\theta_n \quad (29)$$

where $\pi(\theta_n|\mathcal{D}_n, \mathcal{M})$ is the marginal posterior of the model parameters at load cycle n , given by Eq. (23), and $\pi(\mathcal{D}_n^+|\mathcal{D}_n, \theta_n, \mathcal{M})$ is the *prospective likelihood function*, which measures how likely \mathcal{D}_n^+ is predicted by model class \mathcal{M} parameterized by θ_n . The computation of this likelihood function along with the estimation of the multidimensional integral in Eq. (29) are explained in Appendix B.

Note that for this definition, data for the future degradation process \mathcal{D}_n^+ are assumed to be known, which holds true for offline and pseudo-online prognostics analyses, but it could be a hard assumption in a purely online prognostics scenario. In that case, we assume that \mathcal{D}_n^+ can be available through experiments run in similar conditions or by the existence of repeated tests.

3.2. Algorithms for prognostics

The methodology for track degradation prognostics explained above involves the evaluation of multidimensional integrals (recall Eqs. (23) and (27), respectively) which are analytically intractable, except for

some of especial linear cases using Gaussian uncertainties [22]. Approximating methods like Sequential Monte Carlo [22] are required to overcome those integrals by numerical approximation. In prognostics, these approximating methods are typically integrated within algorithms which (1) sequentially estimate the joint PDF of states and model parameters as long as new data are collected, referred to as *particle filtering* algorithms, and (2) extrapolate the updated state estimations into the future in absence of new observations until the EOL is reached. Depending upon implementation both algorithms may not be separable, however, for the sake of clarity, they will be presented here separately in Sections 3.2.1 and 3.2.2, respectively.

3.2.1. Sequential state estimation algorithm

With particle filters, the PDF of the most up to date state of degradation of the track $\pi(z_{0:n}|y_{1:n})$ can be approximated through a set of N weighted *particles* or samples $\{z_{0:n}^{(i)}, \omega_n^{(i)}\}_{i=1}^N$, as:

$$\pi(z_{0:n}|y_{1:n}) \approx \sum_{i=1}^N \omega_n^{(i)} \delta(z_{0:n} - z_{0:n}^{(i)}) \quad (30)$$

where δ is the Dirac delta and $\sum_{i=1}^N \omega_n^{(i)} = 1$. The value of the particle weights $\{\omega_n^{(i)}\}_{i=1}^N$ can be obtained by *sequential importance sampling* [22] each time a new data point arrives, as explained in the pseudocode given as Algorithm 2. Note that a systematic resampling step is proposed in Algorithm 2 (see line 15) to avoid the well-known weight degeneracy problem associated to the sequential importance sampling method [36]. If necessary, a control step on this degeneracy by using the *effective sample size* (ESS) [36] may be incorporated before the resampling step. The interested reader is referred to Orchard and co-workers [35] for further details about sequential state estimation for model-based prognostics.

3.2.2. Particle-filtering based prognostics

Using the particle-filtering approach presented above, the goal is to obtain an up-to-date estimate of the EOL of the track as the PDF of the earliest time when the predicted degradation states reach the failure region $\overline{\mathcal{U}}$. To this end, a particle filter approximation of the predictive Eq. (27) is needed, which can be shown to be obtained as [1]:

$$\pi(z_{n+\ell}|y_{1:n}) \approx \sum_{i=1}^N \omega_n^{(i)} \int_{\mathcal{Z}} \pi(z_{n+1}|z_n^{(i)}) \prod_{t=n+2}^{n+\ell} \pi(z_t|z_{t-1}) dz_{n+1:n+\ell-1} \quad (31)$$

Observe that the last equation cannot be solved analytically, however it can be sampled by drawing one conditional sample sequence $z_{n+1:n+\ell}^{(i)} = \{z_{n+1}^{(i)}, z_{n+2}^{(i)}, \dots, z_{n+\ell}^{(i)}\}$ from each of the N multidimensional integrals in Eq. (31), using the conditional sampling methodology explained in Definition 6. Each of the predicted sequences $z_{n+1:n+\ell}^{(i)}$, $i = 1, \dots, N$ adopts the weight $\omega_n^{(i)}$ of the corresponding updated sample $z_n^{(i)}$; therefore an approximation of Eq. (27) would finally be obtained as:

$$\pi(z_{n+\ell}|y_{1:n}) \approx \sum_{i=1}^N \omega_n^{(i)} \delta(z_{n+\ell} - z_{n+\ell}^{(i)}) \quad (32)$$

where $z_{n+\ell}^{(i)} \in z_{n+1:n+\ell}^{(i)}$, $i = 1 \dots N$. Based on Eq. (32), a particle-filter estimation of the EOL and RUL at time n is obtained as

$$\pi(\text{EOL}_n|y_{1:n}) \approx \sum_{i=1}^N \omega_n^{(i)} \delta(\text{EOL}_n - \text{EOL}_n^{(i)}) \quad (33a)$$

$$\pi(\text{RUL}_n|y_{1:n}) \approx \sum_{i=1}^N \omega_n^{(i)} \delta(\text{RUL}_n - \text{RUL}_n^{(i)}) \quad (33b)$$

In Eqs. (33a) and (33b), $\omega_n^{(i)}$ is the normalised weight of the i th particle at time of prediction n , and $\text{EOL}_n^{(i)}$ and $\text{RUL}_n^{(i)}$ are the particles for EOL_n and RUL_n respectively, which are obtained as

Input: N {number of particles}; $\{z_0^{(i)} = (x_0^{(i)}, \theta_0^{(i)}), \omega_0^{(i)}\}_{i=1}^N$ {particles from prior $\pi(z_0)$ }; $\Sigma_{\xi_0} = \text{diag}(\sigma_{\xi_{0,1}}^2, \sigma_{\xi_{0,2}}^2, \dots, \sigma_{\xi_{0,n_\theta}}^2)$ {initial covariance for artificial dynamics}; $RMAD_j^*$ {target RMAD for j^{th} component of θ }; P_j^* {to control speed of convergence to $RMAD_j^*$ };

Output: $\{z_n^{(i)} = (x_n^{(i)}, \theta_n^{(i)}), \omega_n^{(i)}\}_{i=1}^N$, {updated particles at time n }

Begin ($n \geq 1$): {time n evolves as new data arrive }

1: **for** $i = 1$ to N **do**

2: Sample from Eq. (25): $\theta_n^{(i)} \sim \pi(\cdot | \theta_{n-1}^{(i)})$

3: Sample from Eq. (19a): $x_n^{(i)} \sim \pi(\cdot | x_{n-1}^{(i)}, \theta_n^{(i)})$

4: Set $z_n^{(i)} \leftarrow (x_n^{(i)}, \theta_n^{(i)})$ and $z_{0:n}^{(i)} \leftarrow (x_{0:n}^{(i)}, \theta_{0:n}^{(i)})$

5: Update weights: $\hat{\omega}_n^{(i)} = \omega_{n-1}^{(i)} \underbrace{\pi(y_n | z_n^{(i)})}_{\text{Eq. (19b)}}$

6: **end for**

7: **for** $i = 1$ to N **do**

8: Normalise weights $\omega_n^{(i)} = \frac{\hat{\omega}_n^{(i)}}{\sum_{i=1}^N \hat{\omega}_n^{(i)}}$

9: **end for**

10: **for** $j = 1$ to n_θ **do**

11: Compute $\pi(\theta_{n,j} | y_{1:n}) \approx \sum_{k=1}^N \omega_n^{(k)} \delta(\theta_{n,j} - \theta_{n,j}^{(k)})$

12: Update variance of random walk (use Eq. 26): $\sigma_{\xi_{n,j}}^2 = \sigma_{\xi_{n-1,j}}^2 \left(1 - P_j^* \frac{RMAD(\theta_{n,j}) - RMAD_j^*}{RMAD(\theta_{n,j})}\right)$

13: **end for**

14: Set $\Sigma_{\xi_n} = \text{diag}(\sigma_{\xi_{n,1}}^2, \dots, \sigma_{\xi_{n,n_\theta}}^2)$ in Eq. (25)

15: $\{z_n^{(i)}, \omega_n^{(i)}\}_{i=1}^N \leftarrow \text{resample}\{z_n^{(i)}, \omega_n^{(i)}\}_{i=1}^N$

Algorithm 2. SIS particle filter with parameter adaptation.

Input: $\{z_n^{(i)} = (x_n^{(i)}, \theta_n^{(i)}), \omega_n^{(i)}\}_{i=1}^N$ {updated particles at time n from Algorithm 2}, $\bar{\mathcal{U}} \subset \mathcal{Z}$ {failure domain},

Output: $\{\text{EOL}_n^{(i)}, \omega_n^{(i)}\}_{i=1}^N, \{\text{RUL}_n^{(i)}, \omega_n^{(i)}\}_{i=1}^N$

Begin ($t \geq n$):

- 1: **for** $i = 1, \dots, N$ **do**
- 2: $t \leftarrow n$
- 3: $z_t^{(i)} \leftarrow z_n^{(i)}$
- 4: Evaluate $\mathbb{I}_{(\bar{\mathcal{U}})}(z_t^{(i)})$
- 5: **while** $\mathbb{I}_{(\bar{\mathcal{U}})}(z_t^{(i)}) = 0$ **do**
- 6: Sample from Eq. (25): $\theta_{t+1}^{(i)} \sim \pi(\theta_{t+1}|\theta_t^{(i)})$
- 7: Sample from Eq. (19a): $x_{t+1}^{(i)} \sim \pi(x_{t+1}|x_t^{(i)}, \theta_t^{(i)})$
- 8: $t \leftarrow t + 1$
- 9: $z_t = (x_t^{(i)}, \theta_t^{(i)}) \leftarrow z_{t+1} = (x_{t+1}^{(i)}, \theta_{t+1}^{(i)})$
- 10: **end while**
- 11: $\text{EOL}_n^{(i)} \leftarrow t$
- 12: $\text{RUL}_n^{(i)} = \text{EOL}_n^{(i)} - n$
- 13: **end for**

Algorithm 3. Particle-filter based prognostic algorithm.

$$\text{EOL}_n^{(i)} = \inf\{n + \ell \in \mathbb{N}: \ell \geq 1 \wedge \mathbb{I}_{(\bar{\mathcal{U}})}(z_{n+\ell}^{(i)}) = 1\} \quad (34a)$$

$$\text{RUL}_n^{(i)} = \text{EOL}_n^{(i)} - n \quad (34b)$$

with $\mathbb{I}_{(\bar{\mathcal{U}})}: \mathcal{Z} \rightarrow \{0, 1\}$ being an indicator function that assigns the unity if $(z_{n+\ell} \in \bar{\mathcal{U}})$, and makes zero the rest. An algorithmic description for filtering-based prognostics is provided below as [Algorithm 3](#).

3.3. Prognostics metrics

A prognostic performance evaluation is required to assess the goodness of the EOL/RUL estimates so as to avoid subsequent maintenance decisions based on poor predictions, which may increase system risk. Two main attributes are typically used to evaluate the performance of a prognostic estimation, namely [\[37\]](#): (a) *correctness*, which is related to the prediction accuracy when compared with observed (future) outcomes; and (b) *confidence*, which deals with the uncertainty in the EOL/RUL predictions. Based on these attributes, a prognostic error measure $\mathcal{E}_n: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is proposed here as:

$$\mathcal{E}_n = \frac{\|\mathbb{E}[\text{EOL}_n] - \mathbb{E}[\text{EOL}^*]\|}{\mathbb{E}[\text{EOL}^*]} + \frac{\|\text{std}[\text{EOL}_n] - \text{std}[\text{EOL}^*]\|}{\text{std}[\text{EOL}^*]} \quad (35)$$

where \mathbb{E} denotes the mathematical expectation, std is the standard deviation, and EOL^* represents a benchmark EOL, which is considered as an uncertain variable with known statistics (e.g., EOL^* might be represented by a Gaussian PDF with known mean and standard deviation). Observe from [Eq. \(35\)](#) that \mathcal{E}_n is contributed by the relative discrepancy between the expected values of EOL_n and EOL^* (first term), which accounts for the correctness of the prediction, and the relative difference between the spreads in the PDFs describing the EOL_n and EOL^* (second term), which accounts for the confidence of the prediction.

Remark (Non symmetry). In situations where the computed PDF of EOL_n lacks of smoothness or symmetry, \mathcal{E}_n is preferably formulated using the median (Md), as a measure of location, and the interquartile range (IQR), as a measure of spread, as follows:

$$\mathcal{E}_n = \frac{\|\text{Md}[\text{EOL}_n] - \text{Md}[\text{EOL}^*]\|}{\text{Md}[\text{EOL}^*]} + \frac{\|\text{IQR}[\text{EOL}_n] - \text{IQR}[\text{EOL}^*]\|}{\text{IQR}[\text{EOL}^*]} \quad (36)$$

In addition to quantifying the prognostics performance for a particular model class or a particular algorithm, one may be interested in

assessing and ranking two different model classes, e.g., \mathcal{M}_i and \mathcal{M}_j , according to their relative prognostics performance. To this end, a novel relative prognostics metric is proposed herein based on the ratio between the prospective evidence for each candidate model class (refer to [Definition 9](#)), as follows:

$$\text{RPE}_n = \frac{\pi(\mathcal{D}_{n^+}|\mathcal{D}_n, \mathcal{M}_i)}{\pi(\mathcal{D}_{n^+}|\mathcal{D}_n, \mathcal{M}_j)} \quad (37)$$

where RPE_n stands for the *relative prospective evidence* at time or load cycle n , which provides a measure of the relative anticipation capacity of the two model classes at different time instants during the process, such that $\text{RPE}_n \leq 1$ when $\pi(\mathcal{D}_{n^+}|\mathcal{D}_n, \mathcal{M}_i) \leq \pi(\mathcal{D}_{n^+}|\mathcal{D}_n, \mathcal{M}_j)$.

The reason for proposing a relative performance metric ratio based on evidence is because the computation of the evidence has been shown to automatically enforce a quantitative expression of the Principle of Model Parsimony or Ockham's razor [\[38,39\]](#), so the extremes of over-fitting (at the cost of too much information extracted from data), or under-fitting of the data become naturally penalised. In other words, if the future degradation process \mathcal{D}_{n^+} is predicted equally well by the two models, then the "simpler", i.e., the one which needs less support from the data, would be favoured by the proposed RPE metric. More insight about the proposed prognostics metric will be given below in the context of a case study.

Remark. When dealing with online prognostics, data about the future degradation process \mathcal{D}_{n^+} might be uncertain and known only through a PDF $\pi(\mathcal{D}_{n^+})$. In this case, the RPE metric can be obtained as a mathematical expectation, as

$$\text{RPE}_n = \int \frac{\pi(\mathcal{D}_{n^+}|\mathcal{D}_n, \mathcal{M}_i)}{\pi(\mathcal{D}_{n^+}|\mathcal{D}_n, \mathcal{M}_j)} \pi(\mathcal{D}_{n^+}) d\mathcal{D}_{n^+} \quad (38)$$

The multidimensional integral in [Eq. \(38\)](#) can be numerically evaluated by the Monte Carlo method as

$$\text{RPE}_n \approx \frac{1}{T} \sum_{k=1}^T \frac{\pi(\mathcal{D}_{n^+}^{(k)}|\mathcal{D}_n, \mathcal{M}_i)}{\pi(\mathcal{D}_{n^+}^{(k)}|\mathcal{D}_n, \mathcal{M}_j)} \quad (39)$$

with $\mathcal{D}_{n^+}^{(k)}$ being T sample degradation trajectories drawn from $\pi(\mathcal{D}_{n^+})$.

4. Case study

The knowledge-based prognostics methodology for track settlement presented before is exemplified here using data about permanent axial

Table 1

Permanent vertical strain data used for calculations. Data taken from [23] corresponding to "RTF" test.

| Loading cycles $n, (\times 10^3)$ | 0 | 625 | 1250 | 2500 | 5000 | 10,000 | 20,000 | 50,000 | 75,000 | 100,000 |
|--|---|--------|--------|--------|--------|--------|--------|--------|---------|---------|
| Plastic axial strain ϵ_{1n}^p | 0 | 0.0017 | 0.0045 | 0.0058 | 0.0075 | 0.0087 | 0.0104 | 0.012 | 0.01275 | 0.0133 |

strain in a ballasted railway track taken from the literature [23]. The test, as reported by Aursudkij *et al.* [23], was conducted on the Railway Test Facility (RTF) of the University of Nottingham (UK) [24]. The RTF is housed in a 2.1 [m] (width) \times 4.1 [m] (length) \times 1.9 [m] (depth) concrete pit, and for this test, it was comprised of 0.9 [m] (depth) silt subgrade material, 0.3 [m] (depth) limestone ballast material, and three concrete sleepers. The sleepers were loaded with a sinusoidal load with maximum magnitude of 94 [kN] and 90° phase-lag between each actuator, which simulates a 20 tonnes axle-load train leading to a maximum in-cycle vertical stress $\sigma_{1,max} = 210$ [kPa], and a confining pressure $\sigma_3 = 30$ [kPa] (σ_3 represents the mean pressure due to geostatic stresses in the railway track section). The settlement of the central sleeper was measured using LVDT sensors at a set of non-regularly scheduled load cycles, whereby the permanent axial strain data were obtained by dividing the measured settlements by the ballast section height prior to test (0.3 [m]). A summary of the dataset used for this case study is provided in Table 1. The reader is referred to Aursudkij *et al.* [23] for further information about the experimental setup, and to Brown *et al.* [24] for a detailed description of the Nottingham RTF.

Following the methodology for sequential state estimation presented in Section 3.2, probability-based predictions about the evolution of the permanent axial strain are obtained. The results, which are presented in Fig. 2 for several time instants (load cycles) during the loading process, show a good agreement between the track settlement as measured by the LVDT sensors and that estimated by the particle filter algorithm. For these calculations, the voids ratio of the intact ballast was set to $e_0 = 0.8$, and the values of the critical state parameters shown in Table 2 were adopted to evaluate the underlying constitutive geomechanical model. These critical state parameters have the desirable property of being independent of the state of the material (e.g., voids ratio and stress condition), and, for this case study, they have been obtained through initial fitting tests with the data. In the engineering practice, they can be obtained from routine triaxial tests. Here, the model error term ν defining the state transition equation is assumed to be the same for both the volumetric and deviatoric components, therefore $\Sigma_\nu = \text{diag}(\sigma_\nu, \sigma_\nu)$ in Eq. (19a), where σ_ν is an uncertain parameter that is sequentially updated with the data, as shown further below. The coefficient of variation (c.o.v.) of the measurement error w is set to $\text{c.o.v.}_w = 0.05$, taking it as known, whereby the standard deviation σ_w for the measurement Eq. (19b) can be straightforwardly obtained as $\sigma_w = \text{c.o.v.}_w \|y_n\|$, with y_n being the measurement taken at cycle n . For this case study, the fitting parameters α and β defining the hardening function (recall Eq. (15)) along with the standard deviation of the model error σ_ν , are selected as the uncertain model parameters to be sequentially updated within the augmented state variable $z_n = (x_n, \theta)$, hence $\theta = \{\theta_1, \theta_2, \theta_3\} = \{\alpha, \beta, \sigma_\nu\}$. These

Table 2

Nominal values and prior uncertainty of model parameters and inputs used in calculations.

| Type | Parameter | Value | Prior PDF |
|---------------------------|------------------------|-------------------|---|
| Critical State | Γ | 2.99 | Not applicable |
| | M | 1.9 | Not applicable |
| | λ_{cs} | 0.194 | Not applicable |
| | κ | $7 \cdot 10^{-3}$ | Not applicable |
| Fitting (hardening model) | $\alpha(\theta_1)$ | – | $\mathcal{U}(1, 10)$ |
| | $\beta(\theta_2)$ | – | $\mathcal{U}(0, 1)$ |
| Model error | $\sigma_\nu(\theta_3)$ | – | $\mathcal{U}(5 \cdot 10^{-5}, 5 \cdot 10^{-4})$ |

parameters resulted to be the most sensitive model parameters after a preliminary sensitivity analysis. A uniform PDF is conservatively adopted for the prior PDF of the model parameters $\pi(\theta_j)$, $j = 1, 2, 3$, as specified in Table 2 (see fourth column), as a way of representing our prior state of ignorance about the initial values of such uncertain parameters.

At each prediction step, which corresponds to the loading cycles when settlement data are collected, Algorithm 2 is run by using $N = 5000$ particles to obtain an updated estimation of both the permanent axial strain and the uncertain model parameters. For this algorithm, the scaling variables $RMAD_j^* = 0.1 \cdot RMAD_{0,j}$ and $P_j^* = 0.9$ (recall Eq. (26)) are adopted to control the artificial dynamics of the model parameters $\theta = \{\theta_j\}_{j=1,2,3}$, where $RMAD_{0,j}$ is the relative median absolute deviation of the marginal priors $\pi(\theta_j)$. The diagonal elements $\sigma_{\theta_{0,j}}$ of the covariance matrix Σ_{θ_0} in Eq. (25) are appropriately selected through initial test runs and set to 5% of the 5th–95th inter-percentile range of the marginal priors $\pi(\theta_j)$. To reveal the sequential uncertainty reduction in the updated model parameters, the updated mean estimation of the j th component of θ , as well as its 25% – 75%, 5% – 95% uncertainty bands, are plotted against the loading cycles in Fig. 3 for $j = 1, 2, 3$.

Next, based on the most up-to-date predictions of the permanent axial strain of the track, Algorithm 3 is used for EOL and RUL calculation. To this end, the time-ahead predictions of the track settlement, which are depicted in Fig. 2 using shaded areas to illustrate the prediction uncertainty, are extrapolated forward in time until the permanent axial strain crosses the threshold $\delta\bar{U}$ between the useful domain \mathcal{U} and the failure domain $\bar{\mathcal{U}}$, whereby the PDF of EOL is obtained as a first-passage time distribution problem. For this example, the threshold is set to $\delta\bar{U} = 0.0125$ [dimensionless], thus, the useful domain is defined as the subspace $\mathcal{U} = \{(\epsilon_v^p, \epsilon_s^p) \in \mathbb{R}^2: \epsilon_v^p = \epsilon_s^p + 1/3\epsilon_s^p \in [0, 1.25 \cdot 10^{-2}] \subset \mathbb{R}\}$. To speed up the EOL calculation, which requires the simulation of $N = 5000$ particle trajectories thousands (or even millions) of loading cycles ahead until reaching the failure domain, the "natural" load cycles are converted to duty cycles by dividing the load cycles by a fixed factor $DC \in \mathbb{R} > 1$ [40]. Here, a duty cycle is defined as an appropriate fixed amount of load cycles where a sensible amount of settlement is accumulated. After some pilot tests, $DC = 50$ was revealed as a suitable value for this example.

Fig. 4a shows the sequence of estimated PDFs of EOL for the different load cycles when track settlement data are available. By comparing between consecutive PDFs in Fig. 4a, one can observe that the mean EOL prediction gradually tends to converge to the reference mean value for EOL* (blue marker in Fig. 4a), and the prediction uncertainty (i.e., the spread of the PDFs) tends to decrease, as a consequence of the sequential updating of the model as long as new data are collected. The gradual increase in the prediction accuracy is also revealed in Fig. 4b, where the mean RUL estimations are plotted against time of prediction using an accuracy cone type of representation [41]. Two cones of accuracy at 10% and 20% of the reference RUL, denoted here as RUL* and obtained as $RUL^* = EOL^* - n$, are included to visually interpret the prognostics performance in regards to prediction accuracy. Observe that the expected RUL prediction is appreciably inaccurate for the initial 5000 load cycles (values out of accuracy cones), which suggests that a training stage is required by the model to learn from the data. This learning period is also manifested in Fig. 4c, where the proposed prognostics error \mathcal{E}_n in Eq. (36) is plotted against loading cycles assuming a Gaussian PDF centred at 75,000 and 10% c.o.v. for EOL*.

Finally, the prognostics performance of the track geometry degradation model proposed in Section 2 is compared with the

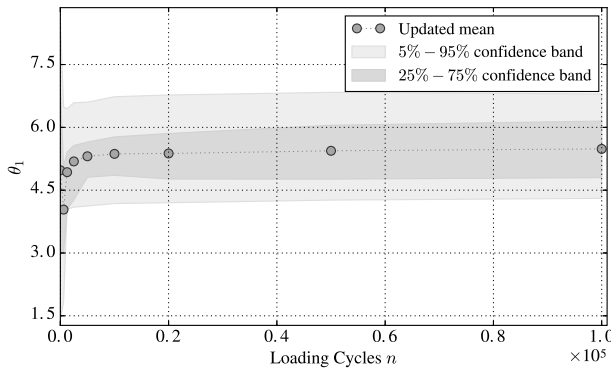
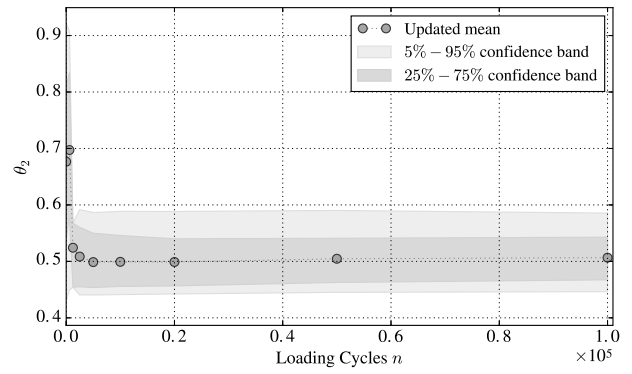
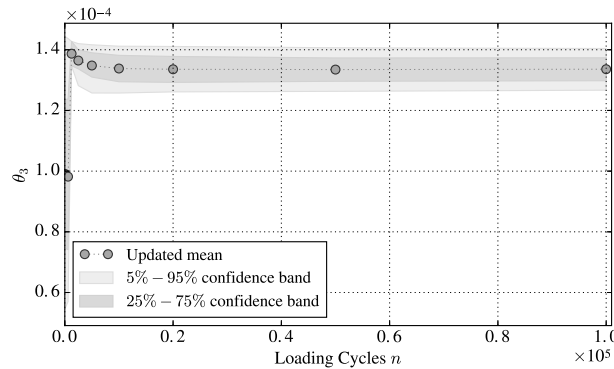
(a) Sequential updating of α parameter(b) Sequential updating of β parameter(c) Sequential updating of model error parameter σ_v

Fig. 3. Trace of the mean values of model parameters $\theta = \{\alpha, \beta, \sigma_v\}$ against load cycles. Shaded areas represent the 25% – 75% (darker color) and 5% – 95% probability bands, respectively.

performance of a well-known class of phenomenological models for track settlement [30,42,43], given by the logarithmic model $\epsilon_n^p = A + B \log n$, where A and B are fitting parameters. For comparison purposes, this model is referred to as *model class* M_0 , while the one proposed in Section 2 is denoted as *model class* M_1 . A discrete-time state-space representation of model class M_0 is obtained and subsequently embedded within the SIS particle filtering algorithm (refer to Algorithm 2), following the methodology in Section 3.1. For model class M_0 , the basic functions \mathbf{f}_n and \mathbf{g}_n defining the hidden Markov model are given by $\mathbf{f}_n = \epsilon_n^p|_{n-1} + B/n$ and $\mathbf{g}_n = \epsilon_n^p|_n$, respectively. The same uncertainty structure (i.e., modelling and measurement errors) that is assumed for model M_1 is also assumed for M_0 . The assessment of the two model classes is carried out by computing the RPE_n metric proposed in Section 3.3. The results are shown in Fig. 5 for the load cycles where measurements are available. In view of Fig. 5, the physics-based model class M_1 proposed in this paper is revealed to be the one with the larger and earlier anticipation capability ($RPE_n > 1$), since it provides larger RPE_n values in the earlier stages of the process. More insight and discussion about prognostics performance is provided in the next section.

5. Discussion

5.1. On the case study results

The proposed methodology for knowledge-based prognostics for railway track degradation has been exemplified using the case study presented in the previous section based on data from the Nottingham Railway Test Facility. As a first interpretation of the results in Fig. 2, the

proposed methodology is able to accurately anticipate the evolution of the plastic vertical strain of the track with quantified uncertainty after an initial learning stage using limited data. The length of such a learning stage is revealed in Fig. 3, where the updated mean of the uncertain model parameters is shown to reach an asymptotic behaviour after the first 5000 load cycles (about 7% of the reference EOL*), which denotes that after this learning period the model has extracted enough information from the data to make accurate predictions. This is in agreement with the results shown in Fig. 4b for the RUL estimations, and also with those in Fig. 4c, where $n = 5000$ is revealed as the time when the prognostics error measure \mathcal{E}_n starts a slower and gradual decrease; i.e., the predictions become increasingly accurate for cycles $n > 5000$, therefore making a decision based on the predicted EOL_n is recommended from that time. Note also in Fig. 4c that the overall prognostics error \mathcal{E}_n is mostly contributed by the uncertainty in the predictions (second term in Eq. (35)), while the error term accounting for the average prognostics accuracy becomes almost negligible after the aforementioned learning period. This early correctness of the prognostics results is a consequence of having adopted a physics-based model for track deterioration instead of a data-based approach. At this point, it is worth mentioning that the learning period required by the model, and consequently, the actual anticipation capacity of the prognostics algorithm, will certainly depend on the early availability and quality of the condition monitoring data, but also on the prior uncertainty of model parameters and their relative “importance” in relation to the overall model response [44]. This model parameter importance is expected to be lower for physics-based prognostics formulations, like the one proposed here, than for purely data-based approaches, where the model response becomes fully dominated by the

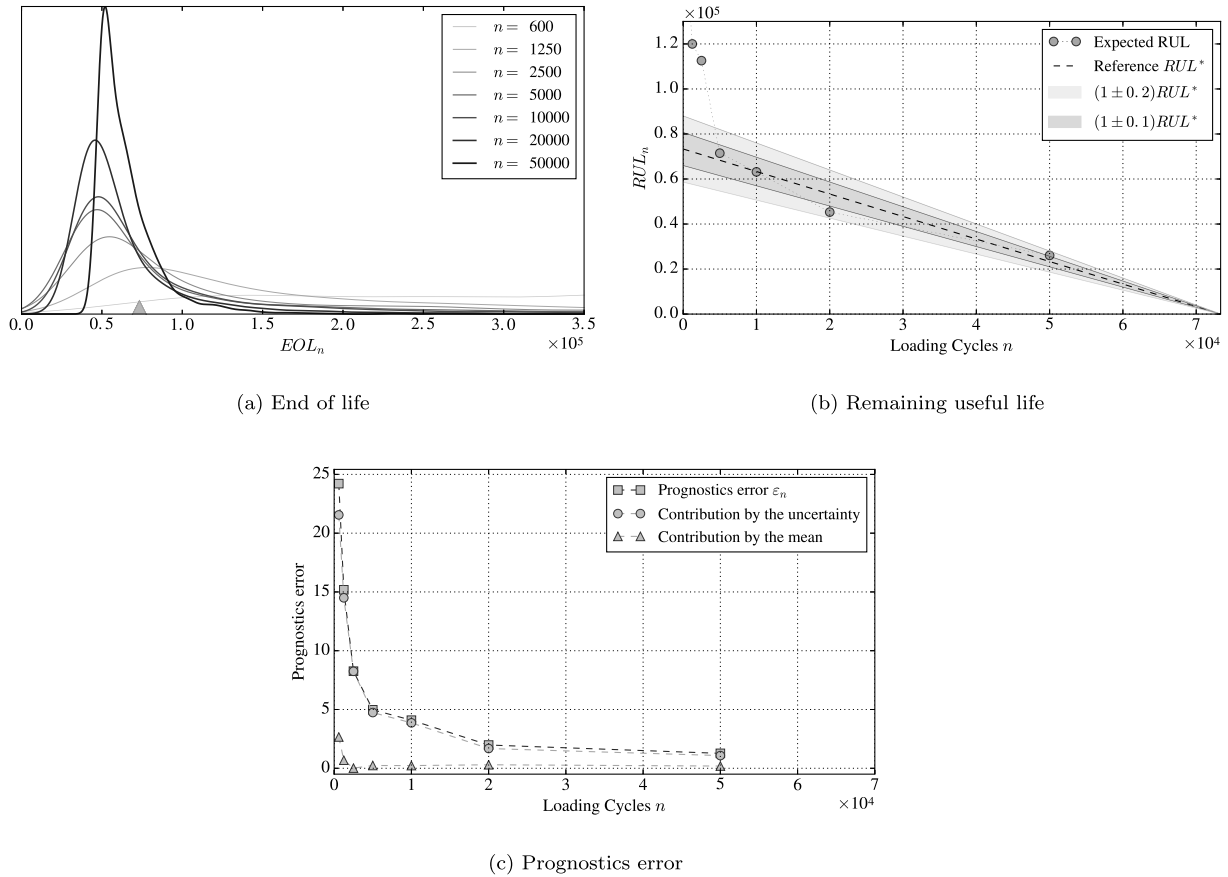


Fig. 4. Results for track degradation prognostics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

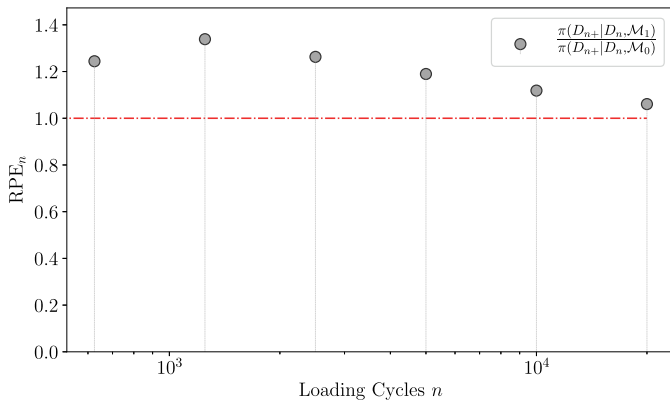


Fig. 5. Relative prospective evidence (RPE_n) obtained for the physics-based model class \mathcal{M}_1 proposed in Section 2 and an empirical logarithmic model class \mathcal{M}_0 .

(uncertain) value of the parameters [45]. This point is reinforced by the values $RPE_n > 1$ in Fig. 5, which reveals that the proposed physics-based model performs better predictions than the data-based logarithmic model used as a benchmark (denoted as \mathcal{M}_0), needing less support from the data to train the fitting parameters. This tendency is more accentuated in the earlier stages of the degradation process due to the initial lack of condition monitoring data. As the process evolves towards the end, both models have extracted enough information from the data to allow them to perform similarly.

In terms of the prediction uncertainty, observe from Figs. 2 and 4 a that the spread in the future predictions, which is initially high due to the high prior uncertainty in the model parameters (especially in the hardening parameters α and β), tends to decrease as the model is sequentially updated with new data, which is a common and desirable feature in any prognostics problem. However, it is also observed that the reduction of this prediction uncertainty does not take place gradually along the full extent of the process but concentrated at the initial stage, approximately coinciding with the learning period of the model where most of the data are available. In this respect, it should be pointed out that the observed uncertainty reduction is a particular consequence of the data collection pattern in this case study, i.e., collected over a set of non-regularly scheduled load cycles mostly concentrated at the beginning of the process, and not a general feature of the proposed physics-based prognostics framework.

Besides the modelling uncertainties discussed above, there is an additional source of uncertainty coming from the prognostics algorithm itself, which is related to the asymptotic approach of the predicted plastic axial strain to the established threshold (as observed in Fig. 2). This asymptotic behaviour makes that many sample predictions need very long simulations to reach the threshold, which leads to poor EOL estimations unless the amount of particles (N in Algorithm 2) is sufficiently large. A suitable solution to overcome this issue is by the adoption of dedicated prognostics algorithms based on high-efficiency sampling techniques, like the one recently proposed by the authors in [46] based on SubSet simulation, or the one developed by Yan *et al.* [47] based on adaptive Lebesgue sampling, among others. This task constitutes a desirable future improvement of this work.

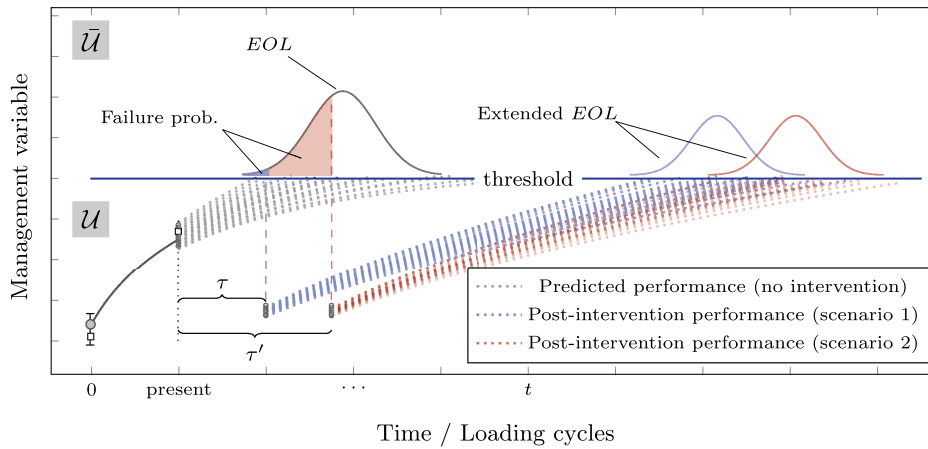


Fig. 6. Illustrative scheme on prognostics-based maintenance. The blue and red dashed lines represent the predicted system behaviour under two alternative intervention scenarios. Observe that delaying an intervention leads to a larger extension of EOL, although it is at the cost of an increasing risk of failure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.2. On the extensibility to track maintenance

As stated in Section 1, the degradation of the railway track geometry results in a key asset management problem for the railway industry with important implications in safety and cost. The potential and also the limitations of the proposed prognostics framework in positively contributing to optimal track asset management is discussed in this section.

First, the authors remark that an apparent limitation of the proposed approach in the context of railway track asset management is that a physical variable such as the plastic strain (or settlement) of the track is adopted instead of more common management variables like the *track geometry index* [48] or the standard deviation (SD) of the settlement [3]. However, the aforementioned management variables can be expressed as functions of the settlement of the track [49] and these can be included within the hidden Markov model in Eq. (18), whereby the proposed filtering-based prognostics methodology is subsequently derived.

Next, in regards to the applicability of the proposed methodology to a real life scenario, it should be pointed out that the physical model described in Section 2 provides the evolution of the permanent settlement of the ballast as a pseudo-plane strain problem. This implies that the railway track is assumed to (approximately) have homogeneous material and geometrical characteristics along the longitudinal direction of loading. For a real world railway track, such assumption would lead to the need for a segmentation of the track into different homogeneous sections where the proposed prognostics framework can be applied. Those sections would constitute predictable components of the track asset for a specific railway line, and other assets related to the track maintenance such as the drainage, switches and crossings might be taken into account to define the length of those sections. In this context, it should be emphasised that the proposed methodology is general and as such, it can be extended to not only other railway assets, but also to other engineering systems and infrastructures. This can be achieved simply by considering the appropriate degradation models and condition monitoring data, and embedding them within a hidden Markov model, as explained in Section 3.1, where the subsequent prognostics methodology is derived. From this standpoint, state-of-the-art methodologies for infrastructure asset management like Petri nets [50,51] can be adopted to perform modelling activities such as optimal maintenance planning and life-cycle cost analyses at system level based on prognostics, which constitutes an immediate next step of this research.

Finally, to illustrate the potential of the proposed methodology for improved maintenance decisions and optimum asset availability, a conceptual scheme has been provided in Fig. 6. In this figure, a degradation process, which is represented using a black solid line, is being

monitored from $n = 0$ up to present time n . At this time, a probabilistic prediction of the future evolution of the process is performed whereby a PDF of the EOL is obtained given a predefined threshold that delimit the end of the safe region \mathcal{U} and the onset of the failure region $\bar{\mathcal{U}}$. Based on this prediction, a decision is made on how much time a corrective maintenance activity can be delayed based on the risk of reaching the failure region during the delay time. To illustrate this decision-making process, two foreseen intervention scenarios are represented in blue and red considering two sample delay times, τ and τ' respectively, with $\tau' > \tau$. Observe that if the intervention is delayed until $n + \tau'$ (scenario 2), a longer lifetime extension is obtained (red PDF) in relation to that obtained for the shorter delay scenario (blue PDF), but it is at the cost of assuming a higher probability of failure (red shaded area in the EOL PDF). Conversely, the shorter delay time scenario (in blue) is more conservative in terms of probability of failure, but it leads to a shorter lifetime extension. This simple example reveals a relevant trade-off between lifetime extension and reliability which is typically faced by infrastructure owners and operators based on experience, which can be rigorously assessed and quantified with the proposed prognostics methodology for infrastructure asset management.

6. Conclusions

A knowledge-based prognostics methodology for railway track asset management has been developed in this paper. The proposed methodology is general but in this paper it was applied to the railway track due to its impact on the safety and the maintenance cost of the overall infrastructure. A geomechanical model for ballast settlement has been provided based on first principles and postulates about the yielding of granular materials. This model is embedded within a particle filtering algorithm for sequential state estimation and prognostics. The algorithm obtains probabilistic estimations of the *remaining useful life* (RUL) of the railway track as the PDF of the time to reach a threshold for tolerable vertical axial strain. A case study has been presented using experimental data for track settlement from the literature to illustrate and discuss the proposed methodology. The results show that the proposed prognostics framework is able to provide reasonably accurate predictions of the RUL after a short model training period. More research effort is needed to exploit the potential of prognostics in the context of optimal track asset management at whole-life whole-system level by integrating prognostics within state-of-art methodologies for infrastructure asset management.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/M023028/1], "Whole-life Cost

Assessment of Novel Material Railway Drainage Systems” along with RSSB who jointly and equally support it, and also the Lloyd’s Register Foundation which partially provides support to this work. RSSB is a rail industry body. Through research, analysis and insight, RSSB supports its members and stakeholders to deliver a safer, more efficient and

sustainable rail system. The Lloyd’s Register Foundation is a charitable foundation in the UK helping to protect life and property by supporting engineering-related education, public engagement and the application of research. The authors would also like to thank Aursudkij, B., McDowell, G. R., & Collop, A. C. [23] for the experimental dataset.

Appendix A. Modelling details

For granular materials like ballast and subballast under a three-dimensional stress state given by the stress tensor σ_{ij} , $i, j = 1, 2, 3$, the following relationships are used to obtain the stress invariants p and q :

$$p = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \quad (\text{A.1a})$$

$$q = \sqrt{\frac{1}{2}((\sigma_{11} - \sigma_{22})^2 + (\sigma_{22} - \sigma_{33})^2 + (\sigma_{33} - \sigma_{11})^2) + 3(\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{13}^2)} \quad (\text{A.1b})$$

The corresponding volumetric and deviatoric strain invariants are given by:

$$\epsilon_v = \frac{1}{3}(\epsilon_{11} + \epsilon_{22} + \epsilon_{33}) \quad (\text{A.2a})$$

$$\epsilon_s = \sqrt{\frac{2}{9}((\epsilon_{11} - \epsilon_{22})^2 + (\epsilon_{22} - \epsilon_{33})^2 + (\epsilon_{33} - \epsilon_{11})^2) + \frac{4}{3}(\epsilon_{12}^2 + \epsilon_{23}^2 + \epsilon_{13}^2)} \quad (\text{A.2b})$$

where ϵ_{ij} is the (i, j) -th element of the strain tensor. Under the assumption of axisymmetric stress state ($\sigma_2 = \sigma_3$ and $\epsilon_2 = \epsilon_3$), the stress and strain invariants simplify to

$$p = \frac{1}{3}(\sigma_1 + 2\sigma_3) \quad (\text{A.3a})$$

$$q = \|\sigma_1 - \sigma_3\| \quad (\text{A.3b})$$

$$\epsilon_v = \epsilon_1 + 2\epsilon_3 \quad (\text{A.3c})$$

$$\epsilon_s = \frac{2}{3}\|\epsilon_1 - \epsilon_3\| \quad (\text{A.3d})$$

From Eqs. (A.3) it follows that:

$$p = q/3 + \sigma_3 \quad (\text{A.4a})$$

$$\epsilon_1 = 1/3\epsilon_v + \epsilon_s \quad (\text{A.4b})$$

The reader is referred to the nomenclature Table A.3 for further information of the geomechanical parameters involved in this work.

Table A.3

Nomenclature table. Nominal values of critical state parameters (Γ , M , λ_{cs} , κ) are given in Table 2.

| | | | |
|-----------------|--|------------|-----------------------------------|
| $d\epsilon_v^p$ | plastic volumetric strain differential | q | deviatoric stress invariant |
| $d\epsilon_s^p$ | plastic distortional strain differential | p | mean stress invariant |
| Γ | critical state model parameter | η | stress-ratio $\eta = \frac{q}{p}$ |
| λ_{cs} | critical state model parameter | p_0 | initial mean stress |
| κ | swelling/recompression constant | σ_1 | vertical stress |
| M | in-plane Poisson ratio | σ_3 | confining pressure |
| α | empirical factor Eq. (15) | e | voids ratio |
| β | empirical factor Eq. (15) | e_0 | initial voids ratio |

Appendix B. Prospective evidence estimation

The *prospective evidence* $\pi(\mathcal{D}_n^+|\mathcal{D}_n, M)$ at time or load cycle n is given by Total Probability Theorem as:

$$\pi(\mathcal{D}_n^+|\mathcal{D}_n, M) = \int_{\Theta} \pi(\mathcal{D}_n^+|\theta_n, \mathcal{D}_n, M) \pi(\theta_n|\mathcal{D}_n, M) d\theta_n \quad (\text{B.1})$$

where $\pi(\mathcal{D}_n^+|\theta_n, \mathcal{D}_n, M)$ is the prospective likelihood function, and $\pi(\theta|\mathcal{D}_n, M)$ is the marginal posterior PDF of model parameters, providing that $\mathcal{D}_n = y_{1:n}$. Since a particle-filter approximation of $\pi(\theta_n|\mathcal{D}_n, M)$ can be available (refer to Eq. (30)), then Eq. (B.1) can be estimated, as:

$$\pi(\mathcal{D}_n^+|\mathcal{D}_n, M) \approx \sum_{i=1}^N \omega_n^{(i)} \pi(\mathcal{D}_n^+|\theta_n^{(i)}, \mathcal{D}_n, M) \quad (\text{B.2})$$

where $\{\theta_n^{(i)}, \omega_n^{(i)}\}_{i=1}^N$ are N particles drawn from $\pi(\theta_n|\mathcal{D}_n, M)$. Assuming that information about the subsequent degradation process is given by $\mathcal{D}_n^+ = \{y_j, y_k, \dots, y_\ell, y_m\}$, the prospective likelihood in Eq. (B.2) is obtained as the probability of predicting \mathcal{D}_n^+ by the stochastic model defined in

Eq. (19a) under the parameterization specified by $\theta_n^{(i)}$ within the model class \mathcal{M} , as follows:

$$\pi(\mathcal{D}_{n^+}|\theta_n^{(i)}, \mathcal{D}_n, \mathcal{M}) = \underbrace{\pi(y_m|y_\ell, \theta_n^{(i)}, \mathcal{M})}_{\text{Eq. (19a)}} \cdots \underbrace{\pi(y_k|y_j, \theta_n^{(i)}, \mathcal{M})}_{\text{Eq. (19a)}} \underbrace{\pi(y_j|y_n, \theta_n^{(i)}, \mathcal{M})}_{\text{Eq. (19a)}} \quad (\text{B.3})$$

where $\{j, k, \dots, \ell, m\} \subset \mathbb{N} > n$ are the times or load cycles where the upcoming data are available. For the general case where \mathcal{D}_{n^+} is available at non-consecutive load cycles (e.g., $k - j \neq 1$), the PDF $\pi(y_k|y_j, \theta_n^{(i)}, \mathcal{M})$ in Eq. (B.3) is obtained as

$$\pi(y_k|y_j, \theta_n^{(i)}, \mathcal{M}) = \int_{\mathcal{X}} \pi(y_k|x_{j:k}, \theta_n^{(i)}, \mathcal{M}) \pi(x_{j:k}|y_j, \mathcal{D}_n, \theta_n^{(i)}, \mathcal{M}) dx_{j:k} \quad (\text{B.4a})$$

$$= \int_{\mathcal{X}} \underbrace{\pi(y_k|x_k, \theta_n^{(i)}, \mathcal{M})}_{\text{Eq. (19b)}} \prod_{t=j}^{k-1} \underbrace{\pi(x_{t+1}|x_t, \theta_n^{(i)}, \mathcal{M})}_{\text{Eq. (19a)}} \underbrace{\pi(x_j|y_j, \mathcal{D}_n, \mathcal{M})}_{\text{Eq. (30)}} dx_{j:k} \quad (\text{B.4b})$$

and the integral in Eq. (B.4) can be numerically estimated by conditional sampling from Eq. (19a), as

$$\pi(y_k|y_j, \theta_n^{(i)}, \mathcal{M}) \approx \frac{1}{N} \sum_{\xi=1}^N \pi(y_k|x_k^{(\xi)}, \theta_n^{(i)}, \mathcal{M}) \quad (\text{B.5})$$

where $x_k^{(\xi)} \in \{x_j^{(\xi)}, x_{j+1}^{(\xi)}, \dots, x_k^{(\xi)}\}$ is the last element of the ξ th conditional sample trajectory drawn from Eq. (19a).

References

- [1] Chiachío J, Chiachío M, Sankararaman S, Saxena A, Goebel K. Prognostics design for structural health management. In: Burgos DAT, Mujica LE, Rodellar J, editors. Emerging design solutions in structural health monitoring systems. IGI Global; 2015. p. 234–73. Ch. 11.
- [2] Esveld C. Modern railway track. 2nd edition The Netherlands: MRT-Productions Zaltbommel; 2001.
- [3] Selig ET, Waters JM. Track geotechnology and substructure management. London: Thomas Telford; 1994.
- [4] Dahlberg T. Some railroad settlement models - a critical review. Proc Inst MechEng Part F 2001;215(4):289–300.
- [5] Soleimanmeigouni I, Ahmadi A. A survey on track geometry degradation modelling. In: Kumar U, Ahmadi A, Verma AK, Varde P, editors. Current trends in reliability, availability, maintainability and safety. Springer; 2016. p. 3–12.
- [6] Higgins C, Liu X. Modeling of track geometry degradation and decisions on safety and maintenance: a literature review and possible future research directions. Proc Inst MechEng Part F 2017;232(5):1385–97.
- [7] Suiker AS, de Borst R. A numerical model for the cyclic deterioration of railway tracks. Int J Numer Methods Eng 2003;57(4):441–70.
- [8] Indraratna B, Thakur PK, Vinod JS, Salim W. Semiempirical cyclic densification model for ballast incorporating particle breakage. Int J Geomech 2012;12(3):260–71.
- [9] Sun Y, Indraratna B, Carter JP, Marchant T, Nimbalkar S. Application of fractional calculus in modelling ballast deformation under cyclic loading. Comput Geotech 2017;82:16–30.
- [10] Hamid A., Gross A.. Track-quality indices and track-degradation models for maintenance-of-way planning, no. 802. 1981.
- [11] Quiroga LM, Schnieder E. Monte carlo simulation of railway track geometry deterioration and restoration. ProcInstMechEng Part O 2012;226(3):274–82.
- [12] Mercier S, Meier-Hirmer C, Roussignol M. Bivariate Gamma wear processes for track geometry modelling, with application to intervention scheduling. Struct Infrastruct Eng 2012;8(4):357–66.
- [13] Vale C, Lurdes SM. Stochastic model for the geometrical rail track degradation process in the portuguese railway northern line. Reliab Eng Syst Saf 2013;116:91–8.
- [14] Prescott D, Andrews J. Investigating railway track asset management using a Markov analysis. ProcInstMechEng Part F 2015;229(4):402–16.
- [15] Bai L, Liu R, Sun Q, Wang F, Xu P. Markov-based model for the prediction of railway track irregularities. Proc Inst MechEng Part F 2015;229(2):150–9.
- [16] Andrade AR, Teixeira PF. Statistical modelling of railway track geometry degradation using hierarchical Bayesian models. Reliab Eng Syst Saf 2015;142:169–83.
- [17] Sankararaman S. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. Mech Syst Signal Process 2015;52:228–47.
- [18] Mishra M, Odelius J, Thaduri A, Nissen A, Rantatalo M. Particle filter-based prognostic approach for railway track geometry. Mech Syst Signal Process 2017;96:226–38.
- [19] Yu HS. Plasticity and geotechnics. Springer Science & Business Media; 2007.
- [20] Schofield A, Wroth P. Critical state soil mechanics. London: McGraw-Hill; 1968.
- [21] Durbin J, Koopman S. Time series analysis by state-space methods. Cambridge University Press; 2003.
- [22] Doucet A, De Freitas N, Gordon N. An introduction to sequential Monte Carlo methods. In: Doucet A, De Freitas N, Gordon N, editors. Sequential Monte Carlo methods in practice. Springer; 2001. p. 3–14.
- [23] Aursudkij B, McDowell GR, Collop AC. Cyclic loading of railway ballast under triaxial conditions and in a railway test facility. Granular Matter 2009;11(6):391–401.
- [24] Brown SF, Brodrick BV, Thom NH, McDowell GR. The nottingham railway test facility, (UK). Proc Inst Civil EngTransp 2007;160(2):59–65.
- [25] Pender M. A model for the behaviour of overconsolidated soil. Geotechnique 1978;28(1):1–25.
- [26] Hill R. The mathematical theory of plasticity. Oxford: University Press; 1950.
- [27] Rowe PW. The stress-dilatancy relation for static equilibrium of an assembly of particles in contact. Proc R Soc London A 1962;269(1339):500–27.
- [28] Roscoe KH, Schofield A, Wroth C. On the yielding of soils. Geotechnique 1958;8(1):22–53.
- [29] Salim W, Indraratna B. A new elastoplastic constitutive model for coarse granular aggregates incorporating particle breakage. Can Geotech J 2004;41(4):657–71.
- [30] Indraratna B, Salim W, Rujikiatkamjorn C. Advanced rail geotechnology–ballasted track. CRC Press; 2011.
- [31] Jaynes E. Papers on probability, statistics and statistical physics. In: Rosenkrantz RD, editor. Kluwer Academic Publishers; 1983.
- [32] Liu J, West M. Combined parameter and state estimation in simulation-based filtering. In: Doucet A, Freitas N, Gordon N, editors. Sequential Monte Carlo methods in practice, statistics for engineering and information science. New York: Springer; 2001. p. 197–223.
- [33] Gordon N, Salmond D, Smith A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE-Proc-F 1993;140:107–13.
- [34] Daigle MJ, Goebel K. Model-based prognostics with concurrent damage progression processes. Syst Man Cybern Syst IEEE Trans 2013;43(3):535–46.
- [35] Orchard ME, Vachtsevanos GJ. A particle-filtering approach for on-line fault diagnosis and failure prognosis. Trans Inst Meas Control 2009;31(3–4):221–46.
- [36] Kong A, Liu JS, Wong WH. Sequential imputations and Bayesian missing data problems. J Am Stat Assoc 1994;89(425):278–88.
- [37] Saxena A, Celaya J, Balaban E, Goebel K, Saha B, Saha S, Schwabacher M. Metrics for evaluating performance of prognostic techniques. Proceedings of the international conference on prognostics and health management. IEEE; 2008. p. 1–17.
- [38] Muto M, Beck JL. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. J Vib Control 2008;14(1–2):7–34.
- [39] Beck J. Bayesian system identification based on probability logic. Struct Control Health Monitor 2010;17(7):825–47.
- [40] Chiachío M, Chiachío J, Rus G, Beck JL. Predicting fatigue damage in composites: a Bayesian framework. Struct Saf 2014;51:57–68.
- [41] Saxena A, Celaya J, Saha B, Saha B, Goebel K. Metrics for offline evaluation of prognostic performance. Int J Progn Health Manage Soc 2010;1(1):4–23.
- [42] Shenton M. Deformation of railway ballast under repeated loading conditions. Railroad track mechanics and technology. Elsevier; 1978. p. 405–25.
- [43] Hettler A.. Bleibende setzungen des schotteroberbaues. Eisenbahntechnische Rundschau 33 (11).
- [44] Chiachío J, Chiachío M, Sankararaman S, Saxena A, Goebel K. Condition-based prediction of time-dependent reliability in composites. Reliab Eng Syst Saf 2015;142:134–47.
- [45] An D, Kim NH, Choi JH. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. Reliab Eng Syst Saf 2015;133:223–36.
- [46] Chiachío M, Chiachío J, Shankararaman S, Goebel K, Andrews J. A new algorithm for prognostics using subset simulation. Reliab Eng Syst Saf 2017;168:189–99.
- [47] Yan W, Zhang B, Wang X, Dou W, Wang J. Lebesgue-sampling-based diagnosis and prognosis for lithium-ion batteries. IEEE Trans Ind Electron 2016;63(3):1804–12.
- [48] Mundrey J. Railway track engineering. Tata McGraw-Hill Education; 2009.
- [49] Chrismers SM. Mechanics-based model to predict ballast-related maintenance timing and costs. Ph.d. thesis. University of Massachusetts Amherst, USA; 1994.
- [50] Andrews J, Prescott D, De Rozières F. A stochastic model for railway track asset management. Reliab Eng Syst Saf 2014;130:76–84.
- [51] Chiachío M, Chiachío J, Prescott D, Andrews J. A new paradigm for uncertain knowledge representation by Plausible Petri nets. Inf Sci 2018;453:323–45.